A randomised sequential procedure to determine the number of factors

Lorenzo Trapani*

Cass Business School, City, University of London

April 21, 2017

Abstract

This paper proposes a procedure to estimate the number of common factors k in a static approximate factor model. The building block of the analysis is the fact that the first k eigenvalues of the covariance matrix of the data diverge, whilst the others stay bounded. On the grounds of this, we propose a test for the null that the i-th eigenvalue diverges, using a randomised test statistic based directly on the estimated eigenvalue. The test only requires minimal assumptions on the data, and no assumptions are required on factors, loadings or idiosyncratic errors. The randomised tests are then employed in a sequential procedure to determine k.

JEL codes: C13, C33.

Keywords: approximate factor models, randomised tests, number of factors.

*Cass Business School, Faculty of Finance, 106 Bunhill Row, London EC1Y 8TZ. Tel.: +44 (0) 207 0405260; email: L.Trapani@city.ac.uk. I wish to thank the Editor (David Ruppert), the Associate Editor and two anonymous Referees for very helpful comments which have greatly improved the paper. This paper originated from a question by Jushan Bai, to whom I am greatly indebted, also for further feedback. I also wish to thank participants in the New York Camp Econometrics X (Syracuse University, April 10-12, 2015), the 21st International Panel Data conference (Central European University, Budapest, 29-30 June, 2015), the Econometric Society World Congress 2015 (Montreal, 17-21 August, 2015), and the Department of Economics at the University of Hull Seminar Series.

1 Introduction

This paper proposes a procedure to determine the number of factors in a static approximate factor model, viz.

$$X_{i,t} = \phi_i' F_t + u_{i,t}, \quad 1 \le i \le N, 1 \le t \le T, \tag{1}$$

where ϕ_i and F_t are column vectors of finite dimension k.

Starting from the seminal contribution by Chamberlain and Rothschild (1983), inference on (1) has been the subject of several studies. In recent years, many contributions have focused on the case of panel data, where both N and T are large - see, *inter alia*, the review of Bai and Ng (2008). The first step in the analysis of (1) is, arguably, the determination of the number of common factors, k. To this end, the literature has developed numerous techniques, which are usually based on a well-established fact: the first k eigenvalues of the covariance matrix of the $X_{i,t}$s diverge to infinity whereas the other ones stay bounded. Two main approaches have been developed. The first one is based on finding a threshold for the eigenvalues of the covariance matrix of the $X_{i,t}$s, which can be used to decide which eigenvalues are finite and which ones are not; the information criteria proposed by Bai and Ng (2002) belong in this category. The second possible approach is based on computing the ratio of adjacent eigenvalues, again exploiting the fact that such ratio eventually diverges: this is the rationale employed by Onatski (2009, 2012) and Ahn and Horenstein (2013), *inter alia*. Neither approach is free from problems. Typically, eigenvalue thresholding requires the choice of a penalty function, as is customary in the context of information criteria (see Bai and Ng, 2002). However, such choice is not unique, which is bound to affect at least the finite sample properties of the estimated k; note however that, building on an idea in Hallin and Liska (2007), Alessi, Barigozzi and Capasso (2010) propose a robust, data-driven methodology to tune the choice of the penalty function which works very well in simulations. Moreover, existing techniques also require comparing the goodness of fit of different versions of (1), for $1 \le k \le k_{\max}$; results seem to be rather sensitive to the specification of the upper bound $k_{\max}$ for at least some of the proposed approaches (see the Monte Carlo evidence in Ahn and Horenstein, 2013). On the other hand, the use of the eigenvalues ratio ameliorates such arbitrariness; nonetheless, existing contributions make extensive use of (large) random matrix theory (see Bai, 1999, for a complete and insightful review), which requires several constraints on the form and amount of serial and cross sectional dependence. Moreover, a standard requirement is that the sample sizes N and T are not too different from each other, usually assuming that, as $\min\{N,T\} \to \infty$, $\frac{N}{T} \to c \in (0,\infty)$.

Hypotheses of interest and testing approach

Let $X_t \equiv \left[X_{1,t},...,X_{N,t}\right]'$: in this paper, we propose a test for the null that the p-th eigenvalue (say $\lambda^{(p)}$) of $E\left(X_t X_t'\right)$ diverges to positive infinity, versus the alternative that it is bounded:

$$\begin{cases} H_0 : \lambda^{(p)} \to \infty \\ H_A : \lambda^{(p)} < \infty \end{cases} . \tag{2}$$

The tests are then employed as part of a sequential procedure to determine k.

From a methodological point of view, the test statistic employed in this paper mimics the behaviour of $\lambda^{(p)}$ - that is, it diverges to positive infinity under H_0. Owing to such lack of randomness under the null, we base our tests on randomising the test statistic. This approach is not new, *per se*, in the literature: the original idea dates back to Pearson (1950), and it has been recently introduced in econometrics - see e.g. Corradi and Swanson (2006). In particular, in this paper we follow the approach used in Corradi and Swanson (2006), where randomisation is employed in conjunction with sample conditioning: randomness is added to the basic statistic, and then the asymptotics is derived conditional on the sample, showing its validity for all samples, save for a zero measure set. Therefore, as explained in Corradi and Swanson (2006), the notion of size is different from the standard one: classically, the level α of a test means that, if a researcher applies the test B times and the null is valid, then (s)he will reject the null with frequency α - that is, (s)he will be wrong αB times. Conversely, in this context α means that out of J researchers who apply the test, αJ of them will reject the null when this is true. Still, in our paper we obtain a test statistic which, for a given level α, rejects the null with probability α (when true), and with probability 1 (when false).

Although randomisation is still not widely employed in econometrics, it has several advantages in this context. To begin with, it affords to actually test for the existence of common factors, rather than being a diagnostics. Also, no restrictions are required on the relative rate of divergence of N and T as they pass to infinity: the number of factors can be estimated for any values of N and T as long as $\min\{N,T\} \to \infty$. This feature makes the test applicable in virtually any context, and it may be helpful in several applications where one dimension is much larger than the other - examples include such diverse fields as accounting (where data are often recorded on an annual basis and are available for many companies, but for a limited number of years), finance (where e.g. data on hedge funds performance are available for thousands of funds, which are live for a relatively short span), microeconometrics with firm level data, marketing studies (where revealed preferences are often recorded over a limited period of time for many consumers), and genomics (where usually thousands of genomes are observed for tens of patients). Finally, it is important to note that the approach adopted in this paper relies on a well-

known asymptotic representation result: on account of (1), the process $X_{i,t}$ is decomposed into two subspaces, the one spanned by (eigenvectors associated with) the spiked eigenvalues of the covarianc ematrix of the data, and its orthogonal complement. Thus, the observations decompose into the sum of their projections onto these subspaces: such decomposition always exists and is unique (we refer to the comments in Hallin and Lippi, 2013). As a consequence, the theory in this paper uses assumptions spelt out directly on the observable quantities $X_{i,t}$, rather than on F_t or $u_{i,t}$.

The remainder of the paper is organised as follows. In Section 2, we spell out the main assumptions and derive preliminary results. The test and its properties (null distribution and consistency) are discussed in Section 3; in particular, in Section 3.3, we discuss the sequential procedure to determine k. Section 4 contains a set of simulations to verify the properties of the test for no factor structure, and of the whole procedure to determine k. Section 5 concludes. Extensions to the case of weak factors and proofs are in the Supplemental Material to the paper (Trapani, 2016).

NOTATION We denote the ordinary limits as "$\to$", and use the symbol "$\asymp$" to indicate that two sequences, say $a_{N,T}$ and $b_{N,T}$, have the same order of magnitude, i.e. $a_{N,T} = O_p\left(b_{N,T}\right)$ and $b_{N,T} = O_p\left(a_{N,T}\right)$. We use "a.s." as short-hand for "almost surely", and "$\equiv$" for definitional equality. The notation M (and, where needed, M', M'', etc...) denotes a finite, generic constant that may differ from line to line. Other relevant notation is introduced in the remainder of the paper.

2 Assumptions and preliminary theory

Consider the matrix form of (1)

$$X_t = \Phi F_t + u_t; \tag{3}$$

in (3), $u_t \equiv \left[u_{1,t},...,u_{N,t}\right]'$ and Φ is an $N \times k$ matrix whose i-th row is ϕ_i'. Henceforth, we assume, without loss of generality, that the data have mean zero, and also that common factors and idiosyncratic errors are orthogonal, as is typical in this literature.

Assumption 1. It holds that *(i)* $E\left(X_{i,t}\right) = 0$ for $1 \le i \le N$ and $1 \le t \le T$; *(ii)* $E\left(F_{j,t}u_{i,t}'\right) = 0$ for $1 \le j \le k$, $1 \le i \le N$ and $1 \le t \le T$.

Assumption 1 *(ii)* is actually a consequence of the asymptotic representation in (1). By Assumption 1, $T^{-1}\sum_{t=1}^{T} E\left(X_t X_t'\right) \equiv \Sigma_X = \Phi\Sigma_F\Phi' + \Sigma_u$, having defined $\Sigma_F \equiv T^{-1}\sum_{t=1}^{T} E\left(F_t F_t'\right)$ and $\Sigma_u \equiv T^{-1}\sum_{t=1}^{T} E\left(u_t u_t'\right)$.

The following notation will also be used extensively henceforth: the p-th largest eigenvalue of Σ_X is denoted as $\lambda^{(p)}$; the p-th eigenvalue of $\Phi\Sigma_F\Phi'$ as $\gamma^{(p)}$; and, finally, the p-th eigenvalue of Σ_u as $\omega^{(p)}$.

Assumption 2. It holds that *(i)* $\gamma^{(p)} = m_p N$ for $1 \leq p \leq k$ and some $m_p > 0$; *(ii)* $\omega^{(p)} \leq M$ for all $1 \leq p \leq N$; *(iii)* $N^{-1}\sum_{i=1}^{N}\gamma^{(p)} \leq M$ for all N.

Assumption 2 adds some structure to the spectra of $\Phi\Sigma_F\Phi'$ and Σ_u, and it is similar, in spirit, to Assumptions 4, 5 and 8 in Forni, Giannone, Lippi and Reichlin (2009). As far as the $\omega^{(p)}$s are concerned, we require that they all be finite; however, they do not need to be distinct or bounded away from zero, and some or all of them could indeed be zero.

As far as the non-zero $\gamma^{(p)}$s are concerned, part *(i)* of the assumption requires that they diverge to positive infinity, as $N \to \infty$, at a rate $O(N)$. This assumption is typical of factor analysis: e.g. Bai and Ng (2002) require that, in addition to Σ_F being positive definite, $N^{-1}\Phi'\Phi$ tends to a positive definite matrix, which is tantamount to assuming that $\gamma^{(p)}$ passes to infinity at a rate $O(N)$. Such behaviour of the common factors is often referred to, in the literature, as having "strong" or "pervasive" factors (see Onatski, 2015, in particular Assumption 2 and the discussion thereafter). However, weaker factors also could be considered, where $\gamma^{(p)}$ is allowed to diverge at a rate slower than N. For the time being, and in order to make the presentation of results easier to follow, we focus on the case of strong factors only; in the Supplement (Trapani, 2016), we investigate the more general case of $\gamma^{(p)} = m_p N^{1-\nu_p}$ with $\nu_p \in [0,1)$.

Finally, note that Assumption 2 does not require that the $\lambda^{(p)}$s be distinct, or that the diverging eigenvalues be well-separated, which are typical requirement in this literature (see e.g. Wang and Fan, 2017, and also Forni, Giannone, Lippi and Reichlin, 2009).

The following well-known result characterizes the eigenvalues of Σ_X.

Lemma 1 *Let $c^{(p)}$ be a set of nonnegative finite numbers, which are strictly positive for $p \leq k$. Then, under Assumptions 1 and 2(i)-(ii), it holds that, as $N \to \infty$*

$$\begin{cases} \frac{\lambda^{(p)}}{N} \to c^{(p)} \text{ for } 1 \leq p \leq k \\ \lambda^{(p)} \to c^{(p)} \text{ for } k+1 \leq p \leq N \end{cases}. \tag{4}$$

Further, define

$$\overline{\lambda}_N \equiv \frac{1}{N}\sum_{p=1}^{N}\lambda^{(p)}; \tag{5}$$

under Assumptions 1 and 2, it holds that

$$\begin{cases} \limsup_{N\to\infty}\overline{\lambda}_N = \overline{\lambda}^{\sup} < \infty \\ \liminf_{N\to\infty}\overline{\lambda}_N = \overline{\lambda}^{\inf} > 0 \end{cases}. \tag{6}$$

According to Lemma 1, $\lambda^{(p)}$ either diverges at a rate $O(N)$, or it converges to a finite constant (which may well be equal to zero) according as $p \leq k$ or not. Basically, the behaviour of the eigenvalues of Σ_X as N passes to infinity is the same as that of the eigenvalues of $\Phi\Sigma_F\Phi'$.

2.1 Estimation of $\lambda^{(p)}$

Consider $\widehat{\Sigma}_X \equiv \frac{1}{T}\sum_{t=1}^{T} X_t X_t'$, and let $\widehat{\lambda}^{(p)}$ denote the p-th largest eigenvalue of $\widehat{\Sigma}_X$. In order to derive the asymptotics of $\widehat{\lambda}^{(p)}$, we need the following assumption.

Assumption 3. It holds that *(i)* $E\left|X_{i,t}\right|^{4+\epsilon} \leq M$ for $1 \leq i \leq N$, $1 \leq t \leq T$ and some $\epsilon > 0$; *(ii)* $E\left[\max_{1\leq \tilde{t}\leq T}\left|\sum_{t=1}^{\tilde{t}} X_{h,t}X_{j,t} - E\left(X_{h,t}X_{j,t}\right)\right|^2\right] \leq MT$ for $1 \leq h, j \leq N$.

Assumption 3*(ii)* is a high-level condition which deserves more comments. In essence, it poses a constraint on the amount of serial correlation that one can have in the process $\{X_{h,t}X_{j,t}\}_{t=1}^{T}$ - and therefore, albeit indirectly, in $X_{i,t}$. According to the assumption, $X_{i,t}$ does not need to be independent across t, which is a requirement in "classical" Random Matrix Theory (see Bai, 1999). On the other hand, similar restrictions to Assumption 3*(ii)* are customarily employed in the literature on factor models (see e.g. Bai and Ng, 2002; Forni, Giannone, Lippi and Reichlin, 2009; and Onatski, 2015). However, Assumption 3*(ii)* differs from the assumptions typically made in this literature since it restricts the amount of serial dependence directly in the $X_{i,t}$s, as opposed to considering the unobservable quantities F_t and $u_{i,t}$ (see however Forni, Giannone, Lippi and Reichlin, 2009). In this respect, Assumption 3*(ii)*, on account of its involving observable quantities only, should be easier to understand and verify.

Two examples are reported below to illustrate how Assumption 3*(ii)* can be verified from more primitive conditions.

EXAMPLE 1. Assumption 3*(ii)* holds if the data are independent. Indeed, assuming that $\{X_{h,t}, X_{j,t}\}$ is independent across t, Burkholder's inequality (see Lin and Bai, 2010) and Assumption 3*(i)* yield $E\left|\sum_{t=1}^{T} X_{h,t}X_{j,t}\right|^{2+\epsilon} \leq MT^{1+\epsilon/2}$. Theorem B in Serfling (1970, p. 1231) thus entails that Assumption 3*(ii)* holds.

EXAMPLE 2. To consider more general cases of (weak) dependence, assume that $X_{i,t}$ is a stationary process with the representation $X_{i,t} = f_i\left(\varepsilon_{i,t}, \varepsilon_{i,t-1}, ...\right)$ for some measurable function $f_i : \mathbb{R}^\infty \to \mathbb{R}$ and an *i.i.d.* sequence $\{\varepsilon_{i,t}\}$. We say that $X_{i,t}$ is L_2-NED (Near Epoch Dependent; see Ling, 2007) of size $\varrho_i \geq \frac{3}{2}$ on the basis $\{\varepsilon_{i,t}\}$ if

$$\left\| X_{i,t} - E\left(X_{i,t}|\mathscr{F}_i^{s,t}\right)\right\|_2 \leq c_{i,t}s^{-\varrho_i}, \tag{7}$$

where $\mathscr{F}_i^{s,t}$ is the σ-field generated by $\{\varepsilon_{i,t}, \varepsilon_{i,t-1}, ..., \varepsilon_{i,t-s}\}$, $\|\cdot\|_2$ denotes the L_2-norm and $c_{i,t}$ is a sequence of non-negative numbers. Condition (7) is very popular when considering non-linear trans-

formations, and it holds for a wide variety of processes, including linear processes, ARCH and GARCH processes and data from dynamical systems and Volterra series (see Davidson, 2002, *inter alia*). By Assumption 3*(i)*, it follows that $X_{h,t}X_{j,t}$ is L_2-NED of size $\frac{1}{2}$ (see Example 17.17 in Davidson, 1994, p. 273). Thus, by Theorem 17.5 in Davidson (1994, p. 204), $X_{h,t}X_{j,t}$ is an L_2-mixingale of size $\frac{1}{2}$; hence, Assumption 3*(ii)* follows from McLeish's maximal inequality (McLeish, 1975).

The rate of convergence of $\widehat{\lambda}^{(p)}$ is in the following lemma.

Lemma 2 *Under Assumptions 1 and 3, it holds that*

$$\widehat{\lambda}^{(p)} = \lambda^{(p)} + O_{a.s.}\left[\frac{N}{\sqrt{T}}\left(\ln^{1+\epsilon} N\right)\left(\ln^{\frac{1+\epsilon}{2}} T\right)\right], \tag{8}$$

for $1 \leq p \leq \min\{N,T\}$*, where* $\epsilon > 0$.

Lemma 2 contains a strong rate for $\widehat{\lambda}^{(p)} - \lambda^{(p)}$, and it can be viewed as the sample counterpart to the population result in Lemma 1. The rate is valid for any combination of N and T, and for all estimated eigenvalues. Further, the lemma does not require Assumption 2, and therefore it does not require any assumptions on the $\lambda^{(p)}$s: these do not need to be distinct or (when they diverge) well-separated; some of the eigenvalues may be equal to zero; and the eigenvalues that diverge do not need to do it at any special rate. In essence, the lemma states that eigenvalues are estimated with an error which depends on the dimensions of the dataset, N and T; in light of (8), the estimation error is quite large. It is, however, comparatively small for $1 \leq p \leq k$, since $\lambda^{(p)}$ is of order $O(N)$. Conversely, the estimation error is very large when $k+1 \leq p \leq \min\{N,T\}$, compared to $\lambda^{(p)}$, which is bounded. As shown in Section 3, the rates in (8), whilst not necessarily sharp for all estimated eigenvalues, afford the construction of a test statistic for (2).

Albeit only incidental to the main arguments in the paper, the result in (8) can be compared with related findings in the literature. In the context of "classical" Random Matrix Theory, it has been shown that, under the assumptions that $X_{i,t}$ is *i.i.d.* across i and t and that $\frac{N}{T} \to c \in (0,\infty)$, it holds that $\widehat{\lambda}^{(p)} - \lambda^{(p)} = O_{a.s.}(1)$ - see Bai and Yin (1993). Lemma 2 illustrates what happens in the presence of common factors, which introduce a spiked eigenvalue structure, with some of the $\lambda^{(p)}$ diverging. Using different assumptions (chiefly, $N > T$ and a restriction on the rate at which $\lambda^{(p)}$ passes to infinity), Wang and Fan (2017; see Theorem 3.1) derive the limiting distribution of $\widehat{\lambda}^{(p)}$ for $1 \leq p \leq k$, showing asymptotic normality at a rate $\frac{N}{\sqrt{T}}$. This suggests that the strong rate in (8) should be optimal, modulo the logarithmic terms, at least for $1 \leq p \leq k$; note however that Lemma 2 holds for all eigenvalues, not only for the spiked ones. Similarly, in a different context and with slightly more stringent assumptions

on the eigenvalues, Forni, Giannone, Lippi and Reichlin (2009, see Lemma 2 *(a)*) show that $\widehat{\lambda}^{(p)} - \lambda^{(p)} = O_P\left(\max\left\{1, \frac{N}{\sqrt{T}}\right\}\right)$. Lemma 2 (which is an almost sure result) yields essentially the same rate when N and T are of comparable magnitude.

3 The test

In this section, we define the test statistic and study its asymptotics under the null and the alternative:

$$\begin{cases} H_0 : \lambda^{(p)} = m_p N \\ H_A : \lambda^{(p)} = m_p < \infty \end{cases},$$

for some $0 < m_p < \infty$ and finite.

3.1 The test statistic

Let $\beta \equiv \frac{\ln N}{\ln T}$, and define $\delta \in [0,1)$ such that

$$\delta \begin{cases} > 0 \\ > 1 - \frac{1}{2\beta} \end{cases} \quad \text{according as} \quad \begin{matrix} \beta \leq \frac{1}{2} \\ \beta > \frac{1}{2} \end{matrix} . \tag{9}$$

Finally, consider the following estimator of $\overline{\lambda}_N$

$$\widehat{\overline{\lambda}}_N \equiv \frac{1}{N} \sum_{p=1}^{N} \widehat{\lambda}^{(p)}. \tag{10}$$

We are now ready to introduce the test. Define

$$\varphi^{(p)} \equiv \exp\left\{ N^{-\delta} \frac{\widehat{\lambda}^{(p)}}{\widehat{\overline{\lambda}}_N} \right\}. \tag{11}$$

Under the null that $\lambda^{(p)} = m_p N$, $\varphi^{(p)} \to \infty$ at a rate $\exp\{N^{1-\delta}\}$; conversely, $\varphi^{(p)}$ converges to a finite number under the alternative that $\lambda^{(p)} < \infty$. We now provide a full-fledged explanation of the latter statement. In order to understand the need for rescaling by $N^{-\delta}$, note that under the alternative it is required that $\varphi^{(p)} = o_{a.s.}(1)$. In essence, this follows as long as $N^{-\delta}\widehat{\lambda}^{(p)} = o_{a.s.}(1)$; in turn, this follows if, on the right-hand side of $N^{-\delta}\widehat{\lambda}^{(p)} = N^{-\delta}\lambda^{(p)} + N^{-\delta}\left(\widehat{\lambda}^{(p)} - \lambda^{(p)}\right)$, both terms are $o_{a.s.}(1)$. The term $N^{-\delta}\lambda^{(p)}$ is $o_{a.s.}(1)$ by assumption. As far as $N^{-\delta}\left(\widehat{\lambda}^{(p)} - \lambda^{(p)}\right)$ is concerned, this should also be $o_{a.s.}(1)$. When $\frac{N}{\sqrt{T}} \to 0$, this is immediately implied by Lemma 2. When $\beta > \frac{1}{2}$, we have $N^{-\delta}\frac{N}{\sqrt{T}} \ln^{\frac{1+\epsilon}{2}} T \ln^{1+\epsilon} N$

$= N^{1-\delta} T^{-\frac{1}{2}} \ln^{\frac{1+\epsilon}{2}} T \ln^{1+\epsilon} N = \beta^{1+\epsilon} T^{-\frac{1}{2}+\beta(1-\delta)} \ln^{\frac{3}{2}(1+\epsilon)} T$; on account of (9), it holds that $-\frac{1}{2} + \beta(1-\delta) < 0$, which yields the desired result. Hence, under the alternative, it follows that $\varphi^{(p)} = o_{a.s.}(1)$. Note that, under the null, $N^{-\delta}\widehat{\lambda}^{(p)}$ diverges, given that $N^{-\delta}\lambda^{(p)} = m_p N^{1-\delta}$ and $\delta < 1$ by construction. Finally, we point out that $\widehat{\widehat{\lambda}}_N$ makes the argument of the exponential scale-free; in principle, any statistic that ensures scale invariance may also be used.

Given that $\varphi^{(p)} \to \infty$ under the null, we cannot use it directly and we instead propose a randomised version of it. We present the construction of the test statistic as a four step algorithm.

Step 1 Generate an artificial sample $\left\{\xi_j^{(p)}\right\}_{j=1}^{R}$ as *i.i.d.* $N(0,1)$, and define the sequence $\sqrt{\varphi^{(p)}} \times \xi_j^{(p)}$, $1 \leq j \leq R$;

Step 2 Define the sample $\left\{\zeta_j^{(p)}(u)\right\}_{j=1}^{R}$ as

$$\zeta_j^{(p)}(u) \equiv I\left[\sqrt{\varphi^{(p)}} \times \xi_j^{(p)} \leq u\right], \tag{12}$$

with u extracted from a distribution $F(u)$ with support $U \subset R \setminus \{0\}$;

Step 3 Compute

$$\vartheta^{(p)}(u) \equiv \frac{2}{\sqrt{R}} \sum_{j=1}^{R} \left[\zeta_j^{(p)}(u) - \frac{1}{2}\right]; \tag{13}$$

Step 4 Define the test statistic

$$\Theta^{(p)} \equiv \int_U \left[\vartheta^{(p)}(u)\right]^2 dF(u). \tag{14}$$

We give a heuristic preview of how the test statistic works. Under the null, $\varphi^{(p)}$ passes to infinity, so that the variance of $\sqrt{\varphi^{(p)}} \times \xi_j^{(p)}$ should be ∞; consequently, the *i.i.d.* sequence $\left\{\zeta_j^{(p)}(u)\right\}_{j=1}^{R}$ follows a Bernoulli distribution with $E\left[\zeta_j^{(p)}(u)\right] = \frac{1}{2}$. Therefore, in (13) a CLT should hold whereby, as $R \to \infty$, $\vartheta^{(p)}(u)$ should be $N(0,1)$. Conversely, under the alternative, $\varphi^{(p)}$ should remain finite, and therefore it can be expected that, for any $u \neq 0$, $E\left[\zeta_j^{(p)}(u)\right] \neq \frac{1}{2}$. Thus, in (13), there is a sum of *i.i.d.* random variables with nonzero mean, which diverges to positive infinity at a speed $\sqrt{R}$.

3.2 Asymptotic properties

We now discuss the null distribution and the power versus $H_A : \lambda^{(p)} \leq m_p < \infty$. Henceforth, we frequently employ the following notation: P^* is the probability law of $\left\{\zeta_j^{(p)}(u)\right\}_{j=1}^{R}$ conditional on the sample, and "$\xrightarrow{D^*}$" denotes convergence in distribution according to P^*.

The following theorem characterizes the null distribution of $\Theta^{(p)}$.

Theorem 1 *Let Assumptions 1-3 hold. Then, under* $H_0 : \lambda^{(p)} = m_p N$, *as* $\min\{N, T, R\} \to \infty$ *with*

$$R \exp\left\{-\epsilon N^{1-\delta}\right\} \to 0, \tag{15}$$

for some $0 < \epsilon < \frac{m_p}{\overline{\lambda}_N}$, *it holds that* $\Theta^{(p)} \xrightarrow{D^*} \chi^2_1$ *a.s.-*P^* *conditionally on the sample.*

Theorem 1 states that, under the null, $\Theta^{(p)}$ follows a chi-squared distribution with one degree of freedom; the result holds for all samples, save for a zero measure set, and no restrictions are needed on the relative rate of divergence of N and T as they pass to infinity.

In order for Theorem 1 to hold, it is necessary that $R \to \infty$, which is natural since equation (13) is an application of the CLT; equation (15) provides an upper bound for R.

Define c_α such that, as $\min\{N, T, R\} \to \infty$, it holds that $P\left[\Theta^{(p)} \leq c_\alpha\right] = \alpha$ under H_0. The following theorem states that the test is consistent versus the alternative $H_A : \lambda^{(p)} \leq m_p$.

Theorem 2 *Let Assumptions 1-3 hold. Under* H_A, *as* $\min\{N, T, R\} \to \infty$, *it holds that* $P\left[\Theta^{(p)} > c_\alpha\right] = 1$ *a.s.-*P^* *conditionally on the sample.*

In the proofs of Theorems 1 and 2, we show that $\vartheta^{(p)}(u)$ has a non-centrality parameter asymptotically equal to

$$\frac{2}{\sqrt{R}} \sum_{j=1}^{R} \int_0^{|u|} \frac{1}{\sqrt{2\pi\varphi^{(p)}}} \exp\left\{-\frac{1}{2}\frac{t^2}{\varphi^{(p)}}\right\} dt = \sqrt{\frac{2R}{\pi}} \left[\frac{|u|}{\sqrt{\varphi^{(p)}}} - \frac{1}{6}\widetilde{u}^3\right],$$

where $\widetilde{u} \in \left(0, \frac{|u|}{\sqrt{\varphi^{(p)}}}\right)$. Under the null, this term should go to zero, whence (15). Under the alternative, the term is bounded from below by

$$\sqrt{\frac{2R}{\pi}} \frac{|u|}{\sqrt{\varphi^{(p)}}} \left[1 - \frac{1}{6}\frac{u^2}{\varphi^{(p)}}\right]; \tag{16}$$

this expression has a local maximum at $|u| = \sqrt{2\varphi^{(p)}}$, and if $\delta > 0$, then $\varphi^{(p)}$ converges to 1; these heuristic considerations point towards choosing $u = \pm\sqrt{2}$.

3.3 Determining k

In this section we study how the individual tests for $H_0 : \lambda^{(p)} \to \infty$ can be used, in a sequential procedure, in order to determine the number of common factors. The estimator of k (say $\widehat{k}$) is the output of

the following algorithm:

Step 1 Run the test for $H_0 : \lambda^{(1)} = \infty$ based on $\Theta^{(1)}$. If the null is rejected, set $\widehat{k} = 0$ and stop, otherwise go to the next step.

Step 2 Starting from $p = 1$, run the test for $H_0 : \lambda^{(p+1)} = \infty$ based on $\Theta^{(p+1)}$, constructed using an artificial sample $\left\{\xi_j^{(p+1)}\right\}_{j=1}^{R}$ generated independently of $\left\{\xi_j^{(1)}\right\}_{j=1}^{R}$, ..., $\left\{\xi_j^{(p)}\right\}_{j=1}^{R}$. If the null is rejected, set $\widehat{k} = p$ and stop; otherwise repeat the step until the null is rejected (or until a pre-specified maximum number, say $k_{\max}$, is reached).

As can be expected, in this context a pivotal role is played by the level of the individual tests, α, which should be chosen so that $\widehat{k}$ is a good approximation of k, at least asymptotically.

Theorem 3 *Let Assumptions 1-3 hold, and define the level of each individual test as $\alpha = \alpha(N,T)$. As $\min\{N,R,T\} \to \infty$ under (15), if $k_{\max} \geq k$ and $\alpha(N,T) \to 0$, then it holds that $P\left[\widehat{k} = k\right] = 1$ a.s.-P^* conditionally on the sample.*

Theorem 3 states that $\widehat{k}$ is consistent, as long as the level α of the individual tests is chosen so as to converge to zero: no specific rates are required (see also Kapetanios, 2010). Further, the theorem does not require any special choice of $k_{\max}$: as long as this value is "large enough" (that is, as long as $k_{\max} \geq k$), the theorem holds. It is worth noting that usually the literature uses the Schwert's rule (Schwert, 1989; see also the comments in Bai and Ng, 2002, p. 203), although other choices are also possible. Indeed, simulations show that the estimation procedure is not sensitive to the choice of $k_{\max}$.

4 Simulations

We evaluate the performance of the sequential procedure to determine k, using synthetic data. Data are generated as

$$X_{i,t} = \sum_{j=1}^{k} \lambda_{i,j} F_{j,t} + \sqrt{\theta} u_{i,t}, \tag{17}$$

where $F_{j,t} \sim i.i.d.\ N(0,1)$ for $1 \leq t \leq T$ and $1 \leq j \leq k$; similarly, $\lambda_{i,j} \sim i.i.d.\ N(1,1)$ for $1 \leq i \leq N$ and $1 \leq j \leq k$. The design in (17) is very similar to Bai and Ng (2002) and Ahn and Horenstein (2013). The idiosyncratic error $u_{i,t}$ is generated as

$$u_{i,t} = \sqrt{\frac{1-\rho^2}{1+2bC}} e_{i,t}, \tag{18}$$

$$e_{i,t} = \rho e_{i,t-1} + v_{i,t} + b\left(\sum_{h=\max\{i-C,1\}}^{i-1} v_{h,t} + \sum_{h=i+1}^{\min\{i+C,N\}} v_{h,t}\right). \tag{19}$$

In (19), $v_{i,t} \sim i.i.d.\ N(0,1)$ for $1 \leq i \leq N$. The coefficient ρ is used to introduce serial dependence in the error term $u_{i,t}$; similarly, the component $b\left(\sum_{h=\max\{i-C,1\}}^{i-1} v_{h,t} + \sum_{h=i+1}^{\max\{i+C,N\}} v_{h,t}\right)$ in (19) introduces cross-sectional dependence among the $u_{i,t}$s. By (18), for most of the units it holds that $Var\left(u_{i,t}\right) = 1$; thus, in (17), θ^{-1} represents the signal-to-noise ratio of the common factors.

Data are generated according to three different schemes, which correspond to different levels of serial and cross-sectional correlation:

(a) *i.i.d.* data: corresponding to $\rho = b = C = 0$;

(b) serially dependent, but cross-sectionally uncorrelated data: $\rho = 0.5$, $b = C = 0$;

(c) serially and cross-sectionally correlated data: $\rho = 0.5$, $b = 0.5$ and $C = \max\left\{10, \frac{N}{20}\right\}$.

Case *(c)* is arguably the most interesting (and problematic) one: the presence of strong cross-sectional dependence in the idiosyncratic term $u_{i,t}$ is observationally equivalent to having a weak common factor whose associated eigenvalue diverges at a rate $N^{1-\nu}$ for ν close to 1.

We report experiments for several combinations of $(N,T) \in \{(25,50,100,200) \times (25,50,100,200)\}$. The cases where all, or most, estimators are uniformly good are not reported to save space.

The test statistics $\Theta^{(p)}$ are specified as follows. Based on (9) we set

$$\delta = \begin{cases} 0.01 & \text{according as} \quad \beta \leq \frac{1}{2} \\ 1.01 \times \left(1 - \frac{1}{2\beta}\right) & \qquad\qquad \beta > \frac{1}{2} \end{cases}. \tag{20}$$

The estimated eigenvalues are rescaled by $\widehat{\widehat{\lambda}}_N$ as suggested in (11) when $N \leq T$; this choice also works for $N > T$, but in this case better results are found using

$$\widetilde{\varphi}^{(p)} \equiv \exp\left\{ N^{-\delta} \frac{\widehat{\lambda}^{(p)}}{\widehat{\widehat{\lambda}}_{N,(p)}} \right\}, \tag{21}$$

with $\widehat{\widehat{\lambda}}_{N,(p)} \equiv N^{-1} \sum_{j=p}^{N} \widehat{\lambda}^{(j)}$. We use $u = \pm\sqrt{2}$, chosen with equal weight. Finally, based on Theorem 3, the level of each test should be chosen so as to go to zero as $\min\{N,T\} \to \infty$; we have employed $0.01/\min\{N,T\}$, which is in the spirit of Bonferroni-type approaches. As a general comment, this (conservative) choice may result in overstating rather than understating k, which could be more desirable.

The procedure suggested here is compared against the methodologies suggested in Bai and Ng (2002; referred to as IC1, IC2, PC1, PC2 below), considering also the refinements developed by Alessi, Barigozzi

and Capasso (2010); Onatski (2010; referred to as ON), and Ahn and Horenstein (2013; referred to as ER and GR):

$$
\begin{aligned}
IC1 &= \arg \min_{0\leq k\leq k_{\max}} \left[\ln V(k) + C_0 k \frac{N+T}{NT} \ln\left(\frac{NT}{N+T}\right) \right] \\
IC2 &= \arg \min_{0\leq k\leq k_{\max}} \left[\ln V(k) + C_0 k \frac{N+T}{NT} \ln\left(\min\{N,T\}\right) \right] \\
PC1 &= \arg \min_{0\leq k\leq k_{\max}} \left[V(k) + C_0 \hat{\sigma}^2 k \frac{N+T}{NT} \ln\left(\frac{NT}{N+T}\right) \right] \\
PC2 &= \arg \min_{0\leq k\leq k_{\max}} \left[V(k) + C_0 \hat{\sigma}^2 k \frac{N+T}{NT} \ln\left(\min\{N,T\}\right) \right] \\
ON &= \arg \max_{0\leq k\leq k_{\max}} \left[k | \widehat{\lambda}^{(k)} > \left(1+N^{-1/3}\right) \hat{u} \right] \\
ER &= \arg \max_{0\leq k\leq k_{\max}} \frac{\widehat{\lambda}^{(k)}}{\widehat{\lambda}^{(k+1)}} \\
GR &= \arg \max_{0\leq k\leq k_{\max}} \frac{\ln\left[1+\widehat{\lambda}^{(k)}/v(k)\right]}{\ln\left[1+\widehat{\lambda}^{(k+1)}/v(k+1)\right]}
\end{aligned}
$$

where

$$
V(k) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(X_{i,t} - \widehat{\phi}_i' \widehat{F}_t\right)^2,
$$

with $\widehat{\phi}_i$ and $\widehat{F}_t$ the estimators of ϕ_i and F_t studied in Bai (2003) under exactly k factors. We define $\widehat{\sigma}^2 = V(k_{\max})$, $\widehat{u} = 2.7\widehat{\lambda}^{(k_{\max}+1)} - 1.7\widehat{\lambda}^{(2k_{\max}+1)}$ and $v(k) = \sum_{j=k+1}^{\min\{N,T\}} \widehat{\lambda}^{(j)}$. In their contribution, Alessi, Barigozzi and Capasso (2010) recommend to employ different values of the tuning constant C_0, and to evaluate the estimated number of factors over a whole range of values of C_0, thereby selecting the optimal one, identified as the value which yields a stable estimate of k. We implemented this procedure by searching for the optimal value of C_0 over the grid $[0,13]$, using intervals of width 0.005; results are reported for $PC1$, which was the best performing criterion across all exercises. We do not report the criteria proposed by Bai and Ng (2002) since the results are usually worse than those obtained using the refinement proposed by Alessi, Barigozzi and Capasso (2010); they are anyway reported in the tables in the Supplement (Trapani, 2017).

We consider two different experiments.

Experiment I: (testing for) no factor structure

We start by considering the case $k = 0$ - that is, the case of no factor structure. In this context, scheme *(c)* is particularly interesting, since it allows for the existence of cross sectional correlation among the data, but not due to common factors. This part of the analysis is related to the papers by Castagnetti, Rossi and Trapani (2015a, 2015b), who propose tests for the null of no factor structures (see also the

discussion in Bai, 2009). All the criteria described above are able, in principle, to determine whether $k = 0$; specifically, the ones designed by Ahn and Horenstein (2013) require an initialisation which we base on $\widehat{\lambda}^{(0)} = \max\{N^{-1/2}, T^{-1/2}\}$. As far as our test is concerned, we recommend that the first step in the analysis should be to carry out a test, at level α (we choose $\alpha = 0.05$), for $H_0 : \lambda^{(1)} = \infty$ versus $H_A : \lambda^{(1)} < \infty$; upon rejecting the null, the conclusion can be reached that $k = 0$; conversely, if the null is not rejected, then the procedure described in Section 3.3 should be employed, thus casting the estimation of k into a two-stage procedure. We implement the test with the (very simple) specification $R = 200$, which works well for all cases considered.

[Insert Figure 1 somewhere here]

The results in Figure 1 show that the test proposed here has excellent power for all cases considered, being able to detect whether $k = 0$ or not. Note that in the worst case scenario, corresponding to serial and cross sectional dependence, with $(N,T) = (25,100)$, the power is anyway in the region of 95%. Indeed, when there is cross sectional dependence, the test proposed in this paper clearly dominates all other approaches.

All other criteria also work very well when there is neither serial nor cross sectional dependence - a possible exception, shown in the Supplement (Trapani, 2017), are the criteria developed by Bai and Ng (2002), but the correction by Alessi, Barigozzi and Capasso (2010) dramatically improves their performance, especially when $N \geq 50$. In presence of serial dependence, the performance of other criteria is also very good, at least in moderate to large samples: in particular, the tests developed by Ahn and Horenstein (2013) works extremely well when $\max\{N,T\} \geq 100$, whereas the test developed by Onatski (2010) yields accurate results as long as $T \geq 100$. All criteria, however, systematically overstate the number of factors in presence of cross-sectional dependence, thereby leading to think that there is a factor structure in the data, when in fact this is absent.

As a final note, we tried the same experiment setting $\theta = 2$; results for $\widehat{k}$ are the same as for $\theta = 1$, and thus we do not report them.

Experiment II: determining the number of (strong) common factors

We evaluate the procedure to determine k considering the cases $k = 1,3$ and 5. In the first set of results (Figure 2), we set $\theta = 1$. As far as the implementation of the test is concerned, the test works very well when using $R = 400$ for all cases considered; indeed, unreported experiments show that the less costly choice $R = 200$ also works well, at least for $N \geq 50$.

[Insert Figure 2 somewhere here]

We do not report results corresponding to scheme *(a)* - no serial or cross sectional dependence - since all estimators perform very well. Figure 2 also shows that results are good all across the board when there is only serial dependence; in this case, there is a slight worsening of $IC1$, $IC2$, $PC1$ and $PC2$ (see the Supplement), which is however limited to when either dimension (N or T) is very small - the impact of small T seems more acute in such case, although the refinement proposed by Alessi, Barigozzi and Capasso (2010) makes these criteria as good as the other ones in this case. Conversely, $\widehat{k}$ and both ER and GR perform very well in this case too. Remarkably, $\widehat{k}$ is the best criterion when N is small - see the cases $(N,T) = (25,100)$ and $(50,50)$ - and it $\widehat{k}$ also works well in the opposite case $(N,T) = (200,25)$, although it tends to understate the true number of factors when $k = 5$. However, when $(N,T) = (200,50)$, $\widehat{k}$ becomes very good even when $k = 5$, which seems to suggest that $\widehat{k}$ performs well for a wide spectrum of values of N, especially when $T \geq 50$.

When there is cross sectional dependence, conclusions become more mixed; the only exception is the criteria developed by Bai and Ng (2002; see the tables in the Supplement), which are systematically wrong and tend to overstate k in all possible cases, irrespective of the values of (N,T) and of the true k. However, when tuning the penalty function as suggested in Alessi, Barigozzi and Capasso (2010), even these estimators become very reliable, especially when $T \geq 50$. As far as the other estimators are concerned, there is no clear winner among the methodologies employed: $\widehat{k}$ fares better than the other criteria when N is quite small - see the cases $(N,T) = (25,100)$, and $(50,50)$, especially under cross sectional dependence and $k \leq 3$. When $k = 5$ and T is small, $\widehat{k}$ has a tendency to understate k; indeed, all estimators fare worse as k increases - one interesting exception is the criterion developed by Onatski (2010), whose performance actually improves as k increases (considering especially the cases where $N > T$). Finally, the criteria developed by Ahn and Horenstein (2013) also work well across all cases considered, with the same exceptions detailed above.

We now turn to the case of a weaker factor structure, which we simulate by using the same design as above but with $\theta = 2$.

[Insert Figure 3 somewhere here]

The figure shows that results are far less clear cut in this case, especially when there is cross sectional dependence: no technique dominantly outperforms the other ones. As a general comment, $\widehat{k}$ is better when k is small, but its performance deteriorates when k increases; similar results are found when

considering the techniques developed by Ahn and Horenstein (2013); similarly to the results in Figure 2, however, the estimator developed by Onatski (2009) works better for large values of k. The estimator developed by Alessi, Barigozzi and Capasso (2010) performs also very well, as long as N is sufficiently large (at least for the case of cross sectional dependence). As far as $\widehat{k}$ is concerned, note that it breaks down, for large values of k, when T is small - the case $(N,T) = (200,25)$ is exemplary in this respect. However, when T increases the estimator improves rapidly - see the case $(N,T) = (200,50)$.

5 Conclusions

We develop a procedure to estimate the number of common factors in a stationary panel factor model. As is typical in this literature, we exploit the fact that the first k eigenvalues of the data covariance matrix diverge as $N \to \infty$, whilst the other ones stay bounded. We therefore derive a test statistic, based on sample eigenvalues, which diverges or converges according as the corresponding population eigenvalue is unbounded or finite. Given that, under H_0, the test statistic diverges, we suggest a randomised tests in order to recover standard normal inference. The individual tests are then used as part of a sequential procedure to determine k. Results are derived under minimal assumptions; we show that the estimator of k is robust to a wide variety of data features, including serial and cross sectional dependence, presence of weak factors and several combinations of N and T. A noteworthy feature of the proposed test is that it is very good at determining whether a factor structure does actually exist in the data or not. The setup developed in this paper hinges on a static factor model; however, it would be desirable to consider also dynamic factor models. A possible approach would be based on casting the dynamic factor model into a static one, following the approach proposed by Bai and Ng (2007).

Finally, a word of warning on the meaning of the hypotheses tested for. As is natural to think, the question whether an eigenvalue is infinity or not is clearly ill-posed (see Trzincka, 1986). However, the test proposed here is a test on the divergence rate of estimated eigenvalues: despite the asymptotic characterization of the test, the purpose of the analysis in this paper is to assess the *magnitude* of an eigenvalue, rather than its actual behaviour at infinity, thus allowing a researcher to decide whether the p-th eigenvalue of the covariance matrix of the data is "large enough" so that a model with at least p common factors is a good characterization of the $X_{i,t}$s or not.

References

Ahn, S.C., Horenstein, A.R., 2013. Eigenvalue ratio test for the number of factors. Econometrica, 81, 1203–1227.

Alessi, L., Barigozzi, M., Capasso, M., 2010. Improved penalization for determining the number of factors in approximate factor models. Statistics and Probability Letters, 80, 1806–1813.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica, 70, 191–221.

Bai, J., Ng, S., 2007. Determining the number of primitive shocks in factor models. Journal of Business and Economic Statistics, 25, 52–60.

Bai, J., Ng, S., 2008. Large dimensional factor models. Foundations and Trends in Econometrics, 3, 89–163.

Bai, J., 2009. Panel data models with interactive fixed effects. Econometrica, 77, 1229–1279.

Bai, Z.D., 1999. Methodologies in spectral analysis of large-dimensional random matrices, a review. Statistica Sinica, 9, 611–677.

Bai, Z.D., Yin, Y.Q., 1993. Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. Annals of Probability, 21, 1275–1294.

Castagnetti, C., Rossi, E., Trapani, L., 2015a. Inference on factor structures in heterogeneous panels. Journal of Econometrics, 184, 145–157.

Castagnetti, C., Rossi, E., Trapani, L., 2015b. Testing for no factor structures: on the use of Hausman-type statistics. Economics Letters, 130, 66–68.

Chamberlain, G., Rothschild, M., 1983. Arbitrage, factor structure, and mean-variance analysis on large asset markets. Econometrica, 51, 1281–1304.

Corradi, V., Swanson, N.R., 2006. The effects of data transformation on common cycle, cointegration, and unit root tests: Monte Carlo and a simple test. Journal of Econometrics, 132, 195–229.

Davidson, 1994. *Stochastic Limit Theory*. Oxford: Oxford University Press.

Davidson, J., 2002. Establishing conditions for the functional central limit theorem in nonlinear and semiparametric time series processes. Journal of Econometrics 106, 243–269.

Forni, M., Giannone D., Lippi M., Reichlin L., 2009. Opening the black box: structural factor models with large cross-sections. Econometric Theory, 25, 1319–1347.

Hallin, M., Lippi, M., 2013. Factor models in high–dimensional time series. A time-domain approach. Stochastic Processes and their Applications, 123, 2678–2695

Hallin, M., Liska, R., 2007. Determining the number of factors in the generalized dynamic factor model. Journal of the American Statistical Association, 102, 603–617.

Kapetanios, G., 2010. A testing procedure for determining the number of factors in approximate factor models with large datasets. Journal of Business and Economic Statistics, 28, 397–409.

Lin Z., Bai Z., 2010. *Probability Inequalities*. Science Press, Beijing and Springer-Verlag, Berlin.

Ling, S., 2007. Testing for change points in time series models and limiting theorems for NED sequences. Annals of Statistics 35, 1213–1237.

McLeish, D.L., 1975. A maximal inequality and dependent strong laws. Annals of Probability, 3, 829–839.

Onatski, A., 2009. A formal statistical test for the number of factors in the approximate factor models. Econometrica, 77, 1447–1479.

Onatski, A., 2010. Determining the number of factors from empirical distribution of eigenvalues. Review of Economic and Statistics, 92, 1004–1016.

Onatski, A., 2012. Asymptotics of the principal components estimator of large factor models with weakly influential factors. Journal of Econometrics 168, 244–258.

Onatski, A., 2015. Asymptotic analysis of the squared estimation error in misspecified factor models. Journal of Econometrics, 186, 388–406.

Pearson, E.S., 1950. On questions raised by the combination of tests based on discontinuous distributions. Biometrika, 37, 383–398.

Schwert, W., 1989. Test for unit roots: a Monte Carlo investigation. Journal of Business and Economic Statistics, 7, 147–159.

Serfling, R.J., 1970. Moment inequalities for the maximum cumulative sum. The Annals of Mathematical Statistics, 41, 1227–1234.

Trapani, L., 2017. Supplement to: A randomised sequential procedure to determine the number of factors.

Trzcinka, C., 1986. On the number of factors in the arbitrage pricing model. Journal of Finance, XLI, 347–368.

Wang, W., Fan, J., 2017. Asymptotics of empirical eigen-structure for high dimensional spiked covariance. Forthcoming at the Annals of Statistics.

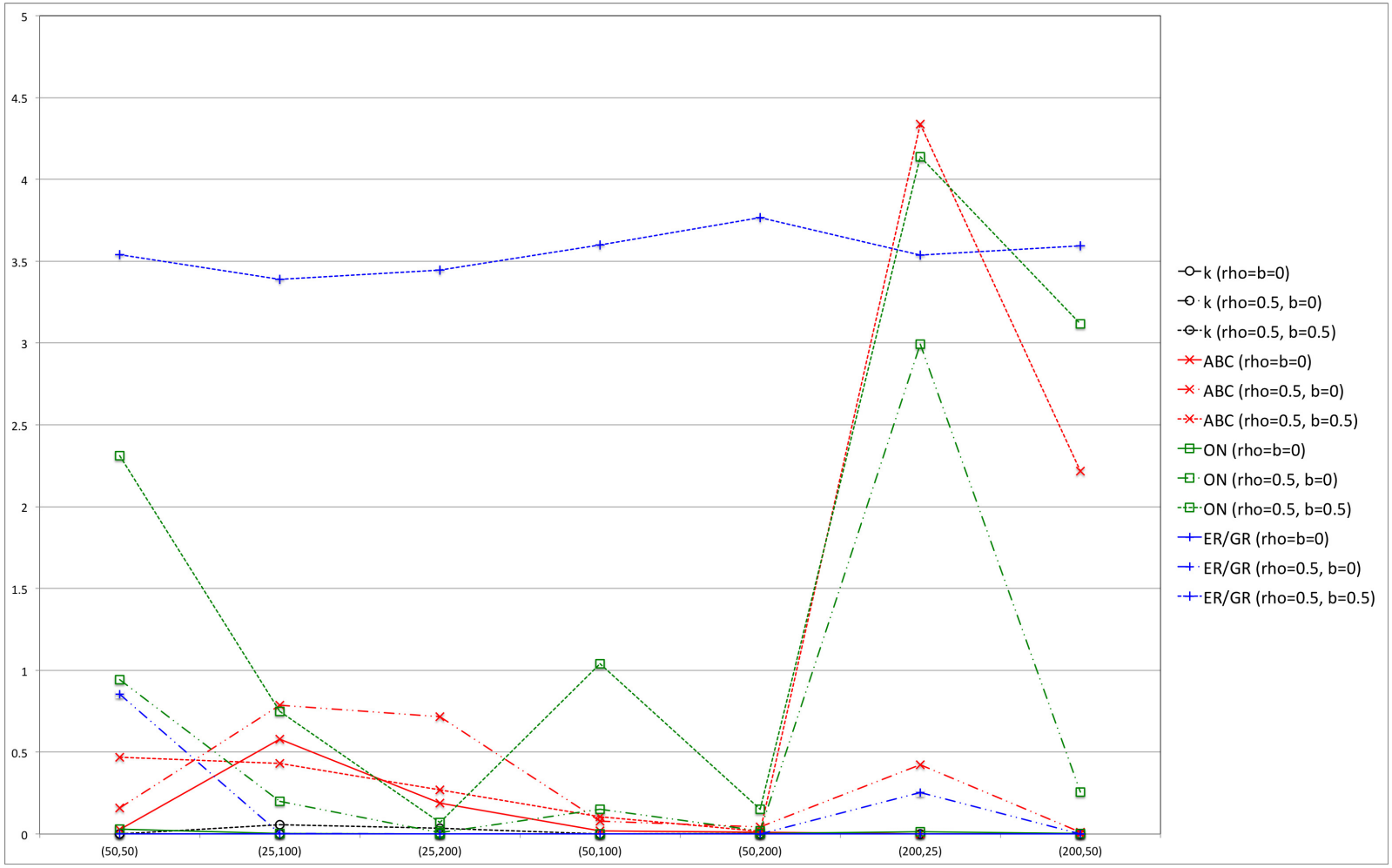

Figure 1: Estimated number of factors when $k = 0$. Data have been generated according to equations (17)-(19), which in this case boil down to $X_{i,t} = \sqrt{\theta} u_{i,t}$. In the first three columns, we set $\theta = 1$. The notation ABC refers to the criteria developed in Alessi, Barigozzi and Capasso (2010); ER/GR means that we report the best between the two criteria.

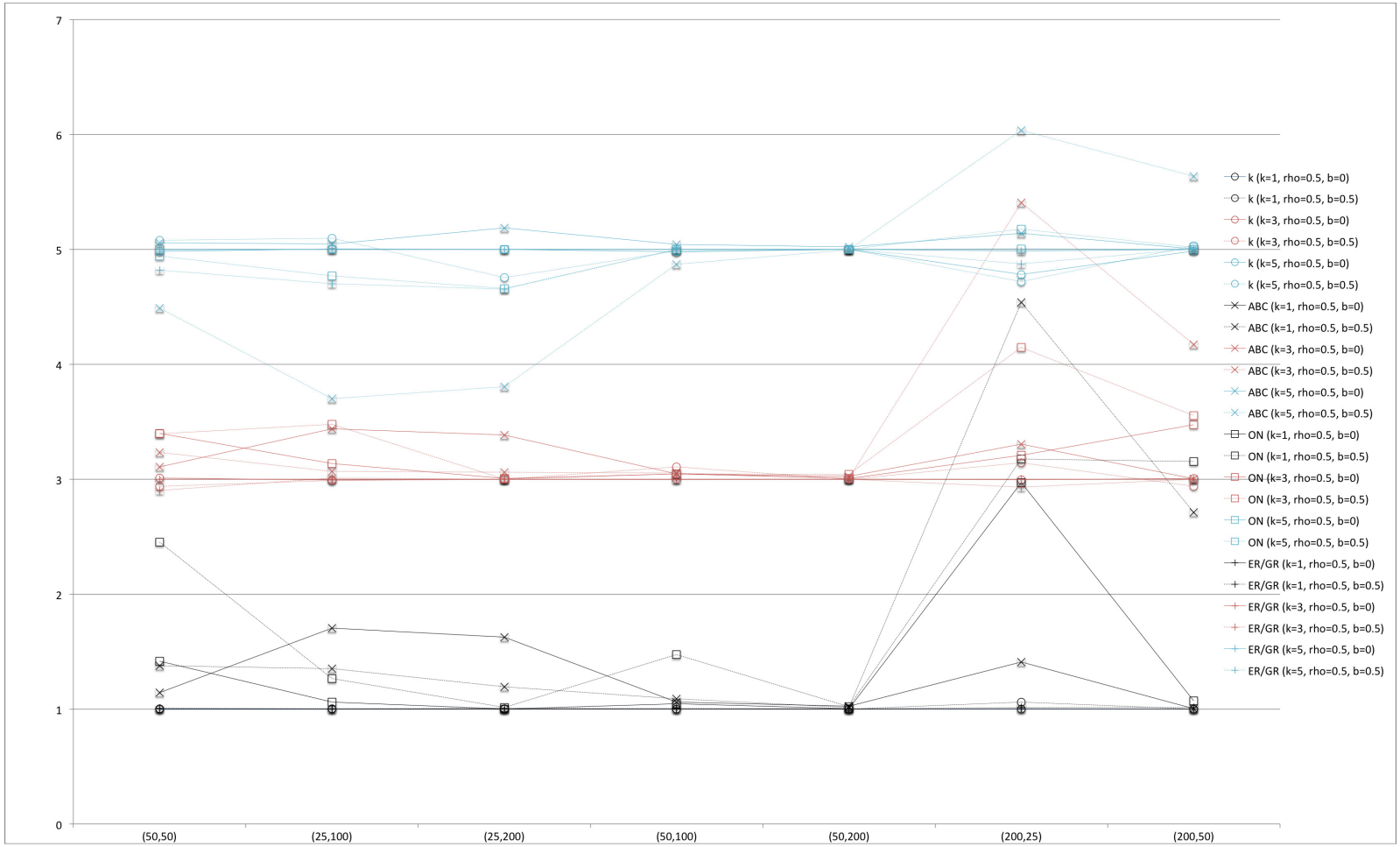

Figure 2: Average number of estimated factors. Data have been generated according to equations (17)-(19) with $\theta = 1$. Only the cases of serially, and serially and cross-sectionally dependent data are reported. As in Figure 1, the notation ABC refers to the criteria developed in Alessi, Barigozzi and Capasso (2010); ER/GR means that we report the best between the two criteria.

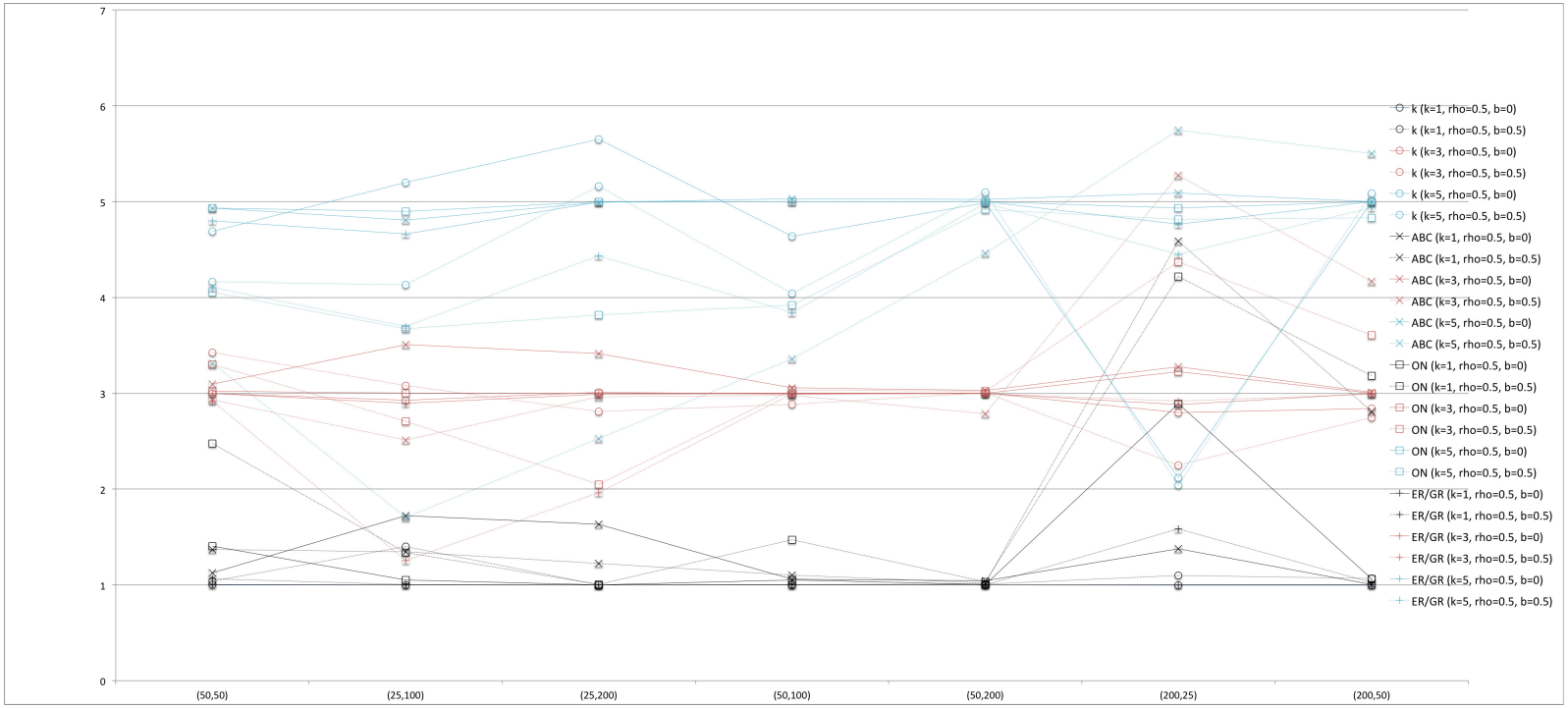

Figure 3: Average number of estimated factors. Data have been generated according to equations (17)-(19) with $\theta = 2$. Only the cases of serially, and serially and cross-sectionally dependent data are reported. As in Figure 1, the notation ABC refers to the criteria developed in Alessi, Barigozzi and Capasso (2010); ER/GR means that we report the best between the two criteria.